*Journal of the Ergonomics Society of Korea*
Vol.19, No.1 2000. 4. 30


# 고개운동에 의한 단순 비언어 의사표현의 비전인식*

# Vision-based recognition of a simple non-verbal intent representation by head movements

유기호**, 노덕수***, 이성철**

## ABSTRACT

In this paper the intent recognition system which recognizes the human's head movements as a simple non-verbal intent representation is presented. The system recognizes five basic intent representations, i.e., strong/weak affirmation, strong/weak negation, and ambiguity by image processing of nodding or shaking movements of head. The vision system for tracking the head movements is composed of CCD camera, image processing board and personal computer. The modified template matching method which replaces the reference image with the searched target image in the previous step is used for the robust tracking of the head movements. For the improvement of the processing speed, the searching is performed in the pyramid representation of the original image. By inspecting the variance of the head movement trajectories, we can recognizes the two basic intent representations - affirmation and negation. Also, by focusing the speed of the head movements, we can see the possibility which recognizes the strength of the intent representation.

Keywords: non-verbal intent representation, head movement, affirmation and negation, vision-based recognition, modified template matching method

* 본 논문의 일부는 1999년도 전북대학교 지원연구비에 의하여 연구되었음
** 전북대학교 기계공학부
주소 : 561-756 전주시 덕진구 덕진동 1가 664-14
전화 : 0652-270-2471
E-mail : yu@moak.chonbuk.ac.kr
*** 전북대학교 대학원 기계공학과

# 1. 서 론

## 1.1. 연구의 배경

최근 인간과 기계/컴퓨터사이의 정보공유의 필요성이 심도 있게 논의되고 있다. 정보공유는 인간과 기계/컴퓨터 상호간의 정보전달 및 인식이 능동적으로 이루어질 때에 가능하며, 이와 같은 기능은 최근 주목을 받고 있는 인간공존형 로봇, 지능 시스템, 재활공학 시스템 등에 필수적으로 요구되는 사항이다. 기계의 동작상태나 컴퓨터의 화면출력에 대한 사용자의 의사표현(반응)을 기계나 컴퓨터가 능동적으로 인식할 수 있다면, 사용자의 요구나 목적에 접근하는 기계동작이나 컴퓨터의 연산·처리를 사용자의 반응을 보면서 단계적으로 선택할 수 있게된다. 이는 인간·기계 인터페이스 시스템에 있어서 정보인식을 인간측에만 의존했던 종래의 일방적이고 수동적인 형태로부터 상호작용 하는 능동적인 형태로의 전환을 가능하게 한다. 또한 기계나 컴퓨터 조작에 관한 사전 정보가 없는 사용자(어린이, 노인을 포함한 비 숙련자)나 조작을 위한 적절한 신체적 수단이 없는 사용자(장애자)에게는 유효한 정보입력 수단으로도 활용될 수 있다.

인간의 의사소통은 언어정보 및 비언어 정보의 조합으로 정보가 전달되고 있는데, 이중 비언어 정보가 교환되는 전체정보량의 6-7할을 차지하고 있다는 사실은 비언어 정보의 중요성을 시사하고 있다(Sugano and Shibuya, 1997). 비언어 정보에는 표정과 제스처를 비롯하여 다양한 방법이 존재하고 있으나, 기본적인 비언어 정보로서 고개의 상하 및 좌우운동에 의한 긍정(만족)과 부정(불만족) 및 애매함(중립적인 상태)을 고려할 수 있다. 만일 기계가, 애매한 상태를 포함하여 긍정과 부정의 강약의 정도를 구분하여 인식할 수 있다면 5형태의 기본적인 의사표현의 인식이 가능하게 된다. 이는 초보적이기는 하나 인간과 기계 상호간의 능동적인 정보인식을 통하여 지능적이며 인간 친화적인 인간·기계 인터페이스 시스템의 구축을 위한 기초연구가 되며, 특히 고개의 상하좌우 운동에 의한 의사표현은 사용언어 및 연령을 초월한 일반성을 가지고 있으므로 폭넓게 적용할 수 있는 가능성도 가지고 있다.

## 1.2. 연구의 내용

본 연구에서는 Figure 1과 같이 의사표현 수단으로서의 고개운동궤적을 비전 시스템으로 기록하고, 운동궤적의 수평축 및 수직 축에 대한 분산과 운동속도에 주목함으로써 운동형태에 따라 강한 긍정(만족), 약한 긍정(약한 만족), 애매함, 약한 부정(약한 불만족), 강한 부정(강한 불만족)의 5형태로 분류할 수 있는 가능성을 보이고 한계점 및 해결방안도 검토한다. 또한 고개운동을 추적하는데 있어서 참조화상을 매 프레임마다 추적된 화상으로 교체함으로써 양호한 실시간 추적성능을 보인 보완된 Template Matching법도 소개한다.

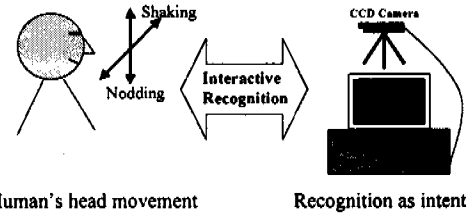



Figure 1. Recognition system of head movements

### 1.3. 관련연구

최근, 감성정보처리 분야에서는 기계/컴퓨터에 의한 감성정보의 표현 및 인식에 관한 연구가 진행되고 있으며, 로보틱스 분야에서는 인간공존형 로봇의 요소기술로서 인간·로봇의 정보공유를 위한 연구가 국내외에서 시도되고 있다. 인간의 표정 인식을 통하여 인간의 심리상태를 이해하려는 시도(Bruce and Young, 1986; Ekman and Friesen, 1977; Kobayashi, et al., 1994; Special issue, 1997; Sugano and Shibuya, 1997)나 제스처 인식을 통하여 인간의 의도를 이해하려는 시도(Fels and Hinton, 1993; Lipscomb, 1991; Rubine, 1991; Special issue, 1997; Yang, et al., 1997, Yu, et al., 1999)등이 그것이다. 이외에도 음성정보나 생체정보를 이용하여 의도 및 심리적 상태를 이해하려는 연구도 시도되고 있다. 이중 본 연구와 관련이 있는 표정 및 제스처 인식은 다양하고 복잡한 표정 및 제스처운동의 실시간 처리, 유연 복잡 구조인 인간의 모델링 수법의 확립 등 해결해야 할 난제를 안고 있다. 반면, 본 연구에서 제안하고 있는 고개운동의 화상처리에 의한 의사표현 인식은 실현 가능성이 높고, 지역, 연령을 초월한 고개운동의 일반성으로 인하여 다수의 불 특정인에 대하여 적용할 수 있는 가능성을 가지고 있음에도 불구하고 거의 연구되고 있지 않는 듯하다.

## 2. 비전인식시스템

### 2.1 실시간 추적알고리즘

피험자의 고개운동을 기록하기 위해서는 안면의 특정부위를 실시간 추적하여 그 궤적을 기록할 필요성이 있다. 실시간 추적을 위한 방법으로는 상관연산을 이용한 Template Matching법이 있는데, 이는 추적하고자 하는 물체의 특징적인 부분을 참조화상(Reference block)으로 하여 탐색영역(Search window)안에서 순차적으로 이동시키면서 상관이 가장 큰 위치를 찾는 방법이다. R을 참조화상, S를 탐색영역으로 하면 상관연산은 다음과 같이 계산된다.

$$C_{i,j} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |R_{m,n} - S_{m+i,n+j}| \quad (1)$$

여기에서 $M,N$은 참조화상의 크기이며, $i,j$를 탐색영역 안에서 변화시키면서 $C_{i,j}$의 최소값을 찾는다. $C_{i,j}$의 값이 최소가 되는 점이 상관이 가장 큰 점이다.

그러나 위와 같은 Template Matching법

은, 인간의 일반적인 고개운동에서 나타나는 안면의 회전 및 거리방향의 이동으로 인한 추적대상 화상의 변형에 대응하기 어려운 문제점을 안고 있다. 본 연구에서는 그 해결책으로 이전의 화상에서 행한 상관연산에서 상관이 가장 높은 화상(추적된 화상)을 참조화상으로 교체하여 다음단계의 처리를 수행하도록 하는 보완된 Template Matching법을 제안한다. Figure 2는 이처럼 참조화상을 처리프레임마다 추적된 화상으로 교체하는 보완된 Template Matching법의 개요를 보이고 있다. 이 방법은 프레임 간격이 짧을수록 즉, 처리속도가 빠를수록 처리 프레임간의 화상변형이 적어 더욱 효과적이다.

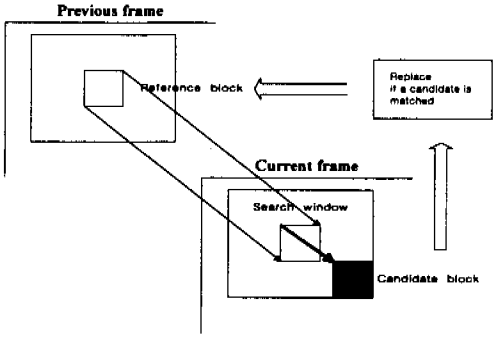



Figure 2. Schematic diagram of the modified template matching method

## 2.2 비전시스템의 구성

실험장치로는 인식대상이 되는 고개운동의 화상신호를 검출하는 CCD 카메라(Panasonic WV-BL200), 카메라로부터 얻어진 신호를 처리하는 화상처리 보드(Data Translation DT2867-LC), 그리고 화상처리 보드로부터 습득한 데이터를 연산 처리하는 컴퓨터(Pentium 100 MHz)로 구성하였다.

## 2.3 Template Matching 모델

탐색을 위한 창(search window)의 영역은 고개운동의 일반적인 범위를 고려하여, 256mm x 256mm의 크기로 하였고, 참조 화상(reference block)으로는 비교적 화상 변형이 적은 코의 하단 부와 인중의 상단 부로 구성되는 부분(42 mm×42 mm)을 택하였다. Figure 3은 실험과정에서 화상데이터를 획득하는 장면을 개략적으로 보여주고 있다. 얼굴 전체부분을 둘러싸는 정사각형이 탐색영역(search window)이고, 코 부분의 작은 정사각형이 참조화상(reference block)이다.

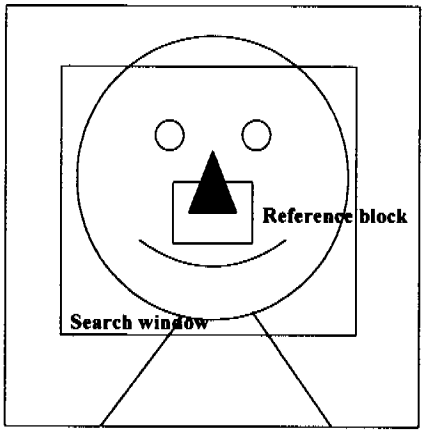



Figure 3. Schematic frame image

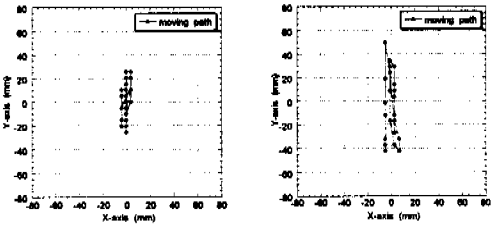
# 3. 고개운동 추적실험

## 3.1 실험방법

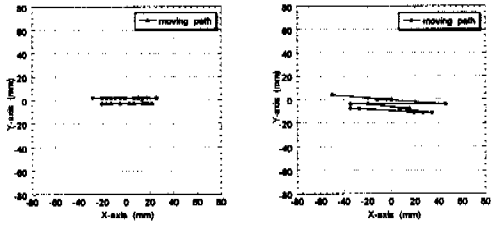
20대 및 30대의 남성을 주로 한 15명으로 피험자 집단을 구성하였다. 피험자를 CCD 카메라로부터 일정한 거리에 앉힌 후, 고개운동에 의한 의사표현을 주문하였다. 피험자에게는 (1)그렇다(긍정), (2)매우 그렇다(강한 긍정) (3)그렇지 않다(약한 부정) (4)전혀 그렇지 않다(강한 부정) (5)잘 모르겠다(애매함)의 순서로 5가지 의사표현을 순차적으로 주문하였다. 한편, 피험자에게는 실험의 목적이나 실험 데이터의 처리방법 등에 대해서 어떠한 정보도 제공하지 않았으며, 자연스러운 분위기에서 평소의 고개운동에 의한 제스처를 유도하였다. 각각의 피험자로부터 고개운동의 궤적을 얻기 위하여 피험자의 코와 인중으로 구성된 참조화상을 설정하고 매초 약 14프레임의 처리속도로 고개운동을 화상처리하였으며, 추적된 참조화상의 중앙 점을 고개운동의 궤적으로 하였다. 또한 획득한 화상 중에서 고개운동이 거의 없는(정지상태에 가까운) 운동전후의 부분을 제외하여 각각의 의사표현에 대한 고개운동 궤적을 얻었다.

## 3.2 실험결과 및 고찰

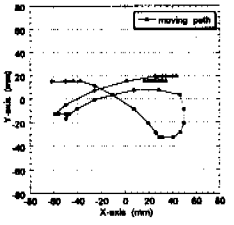
15명으로 구성된 피험자 집단의 고개운동에 의한 의사표현을 화상 처리하여 기록하였다. 일부의 피험자에 대한 반복실험을 포함하여 각각의 의사표현에 대하여 18개의 실험결과를 얻었다. Figure 4는 5형태의 의사표현의 운동궤적의 한 예이다. 긍정의 경우는 수직축 근방에, 부정의 경우는 수평축 근방에 운동궤적이 집중하여 분포하고 있음을 알 수 있다. 또한 동일한 의사표현의 경우라도 강한




(a) Weak affirmation (b) Strong affirmation




(c) Weak negation (d) Strong negation




(e) Ambiguity

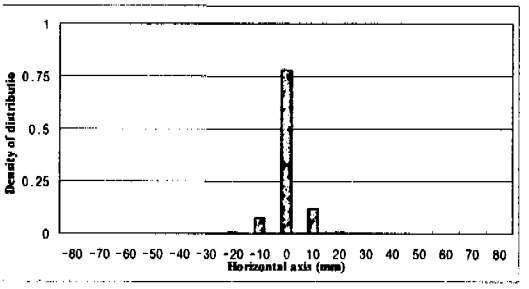
Figure 4. Examples of head movements

의사표현의 경우가 약한 표현에 비하여 분포의 폭이 커지고 있음을 알 수 있다. 애매한 경우는 전반적으로 부정에 가까운 궤적을 보이고 있지만 수직방향의 분산도 적지 않음을

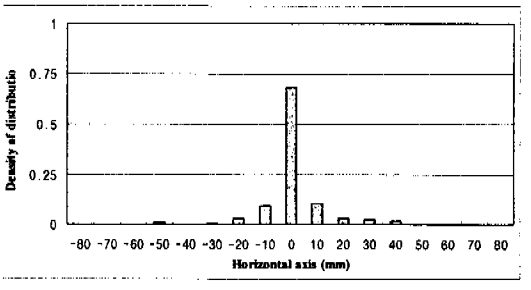
알 수 있다.

이와 같은 고개운동에 의한 비언어 의사표현의 경우 개인차가 존재하기 때문에 피험자 집단의 전체적인 경향을 파악할 필요가 있다. Figure 5는 피험자집단 전체의 운동궤적 밀도를 1로 하였을 때, 각각의 경우에 있어서 분포밀도를 나타내고 있다. 운동궤적의 특성상 긍정의 경우는 수평축에 관하여, 부정의 경우는 수직 축에 관하여, 애매한 경우는 수평, 수직 양 축에 관하여 분포밀도를 나타내었다. 긍정의 경우는 수직 축 근방에, 부정의 경우는 수평축 근방에 각각 궤적이 집중하고 있음을 알 수 있다. 특기할만한 사실은 같은 의사표현의 경우, 강한 표현의 경우가 약한 표현에 비하여 긍정의 경우는 수평축 방향으로, 부정의 경우는 수직 축 방향으로 분산의 정도가 더 커짐을 알 수 있다. 이는 강한 표현의 경우 주 운동방향의 운동폭도 커지지만, 그에 비례해서 주 운동방향에 대한 직각방향으로의 운동폭도 증가하는데 기인한다. 애매한 경우는 궤적이 주로 수평축 근방에 분포하면서도 수직 축으로도 비교적 폭넓게 분포하고 있음을 알 수 있다.

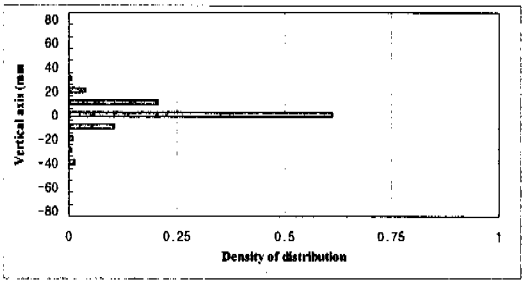
Figure 6은 수직 및 수평축에 대한 분산비와 운동속도의 평균을 매개변수로 하여 5형태의 의사표현을 나타낸 것이다. 수평축은 운동속도의 평균을, 수직 축은 수직 방향의 분산($V_V$)에 대한 수평 방향의 분산($V_H$)의 비율을 나타내고 있다. 분산 비 1을 기준으로 하여 긍정의 경우는 상반부에, 부정 및 애매한 경우는 하반부에 분포되어 있다. 이처럼 불특정 피험자의 고개운동궤적의 분산 비를 통해서 의사표현의 긍정과 부정의 판별이 가능함을 알 수 있다. 그러나 분산 비만으로 강약의 정도까지 판별하는 것은 곤란하며, 운동속도의 정보를 고려할 필요성이 있다. 긍정과




(a) Weak affirmation

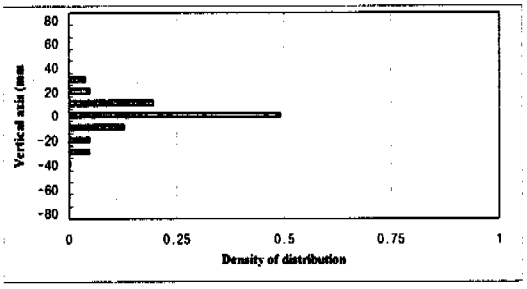



(b) Strong affirmation

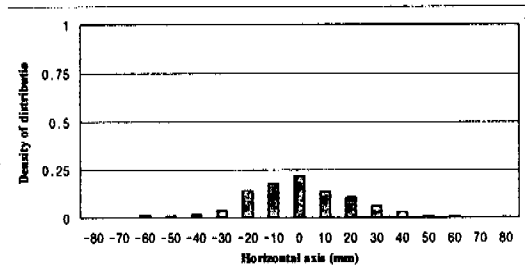


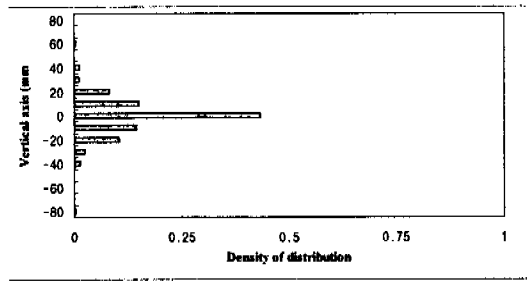

(c) Weak negation

Fig. 5 Density of tracking points (continued)




(d) Strong negation




(e) Ambiguity

Figure 5. Density of tracking points

부정의 경우에 있어서 운동속도로서 강한 표현과 약한 표현을 명확히 구분할 경계치를 찾기는 어렵지만, 전반적으로 강한 표현의 경우가 약한 표현에 비하여 운동속도가 더 큰 것을 알 수 있으며 애매한 경우는 약한 부정보다 운동속도가 더 작음을 알 수 있다. 위에서

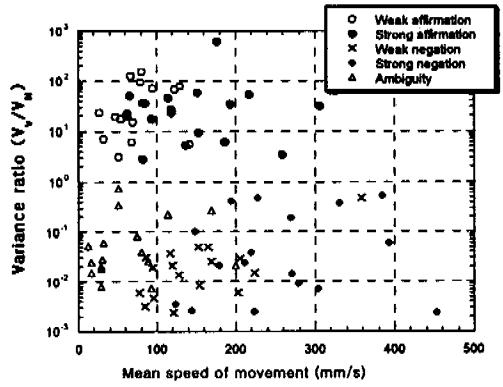



Figure 6. Distribution of 5 basic intent representations

언급한 것처럼 운동속도로서 강약의 표현을 명확히 구분하기 어려운 것은 고개운동의 운동속도에 개인차가 존재하기 때문이다. 그러나 각각의 표현에 있어서 동일 피험자의 고개운동속도는 큰 변화를 보이지 않고 거의 일정한 피험자 고유의 값을 가지고 있음을 반복되는 실험을 통해서 확인할 수 있었다.

Table 1에 각각의 의사표현에 있어서 피험자집단의 분산 비($V_V/V_H$)와 운동속도의 평균, 표준편차 및 매개변수간의 상관계수를 나타내었다. 분산 비와 운동속도 모두 표준편차가 비교적 크게 나타나고 있음을 알 수 있는데, 이는 피험자의 개인차에 의한 영향을 무시할 수 없음을 나타내고 있다. 또한 분산 비의 편차가 운동속도의 편차에 비하여 상대적으로 크게 나타나고 있음을 알 수 있다. 그러나 긍정과 부정이 분산 비 1을 기준으로 명확히 구별되고 있고, 운동속도의 정보까지 고려하면 강약을 구별할 수 있는 가능성도 있어, 본 연구에서 의사표현의 인식을 위하여 채택한 분산 비와 운동속도는 타당한 매개변

수임을 알 수 있다. 분산 비와 운동속도간의 상관관계는 Table 1에 나타난 것처럼 약한 부정의 경우를 제외하고는 대체로 낮은 상관성을 보였다. 약한 부정의 경우 이례적으로 높은 상관관계를 나타내고 있는 것은, Figure 6에서 알 수 있는 것처럼 특정 피험자 1인의 데이터가 집단 평균으로부터 매우 큰 편차를 나타내고 있음에 기인한다. 참고로 매우 큰 편차를 보이고 있는 피험자의 데이터를 제외한 상관계수는 0.551이었다. 전체적으로 낮은 상관관계를 보이면서도 긍정과 부정의 각각의 경우에 있어서 약한 표현의 경우가 강한 표현의 경우에 비하여 상대적으로 높은 상관성을 보이고 있음을 알 수 있다.

Table 1. Statistics of variance ratio and movement speed of subject group

| | Variance ratio ($V_V/V_H$) | | Movement speed (mm/s) | | Correlation coefficient between variance ratio and movement speed |
|---|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation | |
| Weak affirmation | 43.098 | 45.328 | 72.495 | 32.71 | 0.324 |
| Strong affirmation | 60.027 | 140.8 | 144.55 | 68.02 | 0.105 |
| Weak negation | 0.04398 | 0.10727 | 146.88 | 69.553 | 0.732 |
| Strong negation | 0.12588 | 0.18489 | 249.99 | 94.217 | 0.209 |
| Ambiguity | 0.11013 | 0.18404 | 63.234 | 53.849 | 0.126 |

이상과 같은 실험결과의 고찰에 의하면, 불특정 피험자의 고개운동에 의한 의사표현을 화상 처리하여 운동궤적의 분산 비로부터 긍정과 부정의 표현을 판별하는 것이 가능함을 알 수 있다. 그러나 의사 표현의 강약을 판별하는 것은 한계가 있는데, 이는 각각의 피험자의 고개운동속도에 개인차가 존재하므로 인하여 운동속도를 강약의 판단을 위한 보편적인 매개변수로 채택하는데 무리가 따르기 때문이다. 이에 대한 하나의 해결방법으로서는, 각 피험자마다 각각의 의사 표현 상태에 있어서 어느 정도 일정한 고유의 운동속도를 가지고 있는 것에 착안하여 각 피험자의 고개운동에 대한 데이터 베이스를 시스템에 탑재시키거나, 신경회로망 등의 학습알고리즘을 이용하여 피험자의 의사 표현을 학습하면 의사 표현의 강약도 판별이 가능하게 될 것이다.

## 4. 결론 및 전망

고개운동 인식의 화상처리에서, 참조화상을 전 단계에서 탐색된 화상으로 교체함으로써 추적화상의 회전 및 거리방향의 이동에 대응하는 강건한 탐색이 가능하였다. 또한 피라미드 구조에 기초하여 탐색영역의 화소 수를 축소함으로써 처리속도를 개선(초당 14.2프레임)하여 고개운동의 원활한 실시간 추적이 가능하였다.

고개운동의 2차원상의 궤적으로부터 수평축과 수직 축에 대한 분산 비와 고개운동의 평균속도를 매개변수로 하여 강한 긍정, 약한 긍정 및 강한 부정, 약한 부정의 기본적인 의사표현의 인식을 시도하였다. 그 결과, 분산 비로써 긍정과 부정의 인식이 가능하였으나 강약의 판별은 명확한 판단이 어려워 운동속도의 차이에 주목함으로써 강약의 판별 가능성이 있음을 확인하였으며, 신경회로망 등의 학습알고리즘의 적용을 고려할 수 있을 것이다.

본 연구에서는 단순한 의사표현을 인식하는데 있어서 고개운동의 화상처리에만 의존하고

있으나, 일반적으로 단순한 의사표현의 경우에도 고개의 움직임뿐만 아니라 음성을 동시에 활용하는 경향을 고려할 때 음성을 정보처리 하거나 발성시의 입 모양을 화상 처리하여 보완하면 개인차에 의한 추정오차나 주위환경(조명, 소음 등)에 강건하게 대응하는 의사표현 인식시스템의 개발이 가능할 것이다.

## 참고 문헌


Bruce, V. and Young, A.W. "Understanding face recognition," British Journal of Psychology, 77, 305-327, 1986.

Ekman, P. and Friesen, W.V., "Facial Action Coding System," Consulting Psychologist Press, 1977.

Fels, S.S. and Hinton, G.E., "Glove-talk : a neural network interface between a data-glove and a speech synthesizer," IEEE

Trans. Neural Networks, 4, 2-8, Jan, 1993.

Kobayashi, H. et.al., "Real-Time Recognition of Six Basic Facial Expressions," 4th IEEE Int. Workshop on Robot and Human Communication, 179-186, 1994.

Lipscomb, J.S., "A trainable gesture recognizer," Pattern Recognit., 24-9, 895-907, 1991.

Rubine, D.H., "The automatic recognition of gesture," Ph.D. dissertation, Dept. of Computer Science, Carnegie Mellon University, Pittsburg, PA, Dec., 1991.

Special Issue on Face and Gesture Recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.19, no.7, pp.675-785, 1997.

Sugano, S. and Shibuya, K., "Anthropomorphic Robots for Nonverbal Communication," J. of the Robotics Society of Japan, vol.15, no.7, pp.975-978, 1997.

Yang, J., Xu, Y. and Chen, S.S., "Human Action Learning via Hidden Markov Model," IEEE Trans. on Systems and Cybernetics - Part A: Systems and Humans, vol.27, no.1, 1997.

Yu, K.-H., Roh, D.-S., Lee, S.-C., "Recognition of head gestures as a simple intent representations," Proceedings of the 3rd International Workshop on Advanced Mechatronics, 99-103, 1999.


---


**저자 소개**

**◈ 유 기 호**

1988년 전북대학교 정밀기계공학과 졸업. 1990년 동 대학원 기계공학과 졸업(석사). 1994년 일본 Tohoku대학 기계공학과 졸업(박사). 1994년 - 1997년 Tohoku 대학 전

임연구원. 1998년 3월 - 현재 전북대학교 기계공학부 전임강사, 조교수. 관심분야는 메카트로닉스, 맨-머신 인터페이스.

◈ 노 덕 수

1996년 전북대학교 기계공학과 졸업. 1998년 동 대학원 기계공학과 졸업(석사). 현재 군복무중. 관심분야는 화상처리.

◈ 이 성 철

1974년 전북대학교 기계공학과 졸업. 1976년 동 대학원 기계공학과 졸업(석사). 1983년 전남대학교 대학원 기계공학과 박사과정 수료. 1983년 12월 - 1984년 12월 Oklahoma 주립대학 기계항공공학과 객원교수. 1979년 - 현재 전북대학교 기계공학과 전임강사, 조교수, 부교수, 기계공학부 교수. 관심분야는 로보틱스, 기계계측 및 제어.

---

논문접수일 (Date Received): 1999/5/10
논문게재승인일(Date Accepted): 1999/10/3